Analyzing Co-Purchase Network of DVDs in Japanese Online Store

S. Itoh[1,2], S. Takami[3] and T. Iba[4,5]

[1]Graduate School of Media and Governance, Keio University, Endo 5322, Fujisawa, Kanagawa, Japan
[2] Fellow Assistant, Rakuten Institute of Technology Lab., Shinagawa Seaside Rakuten Tower, 4-12-3 Higashi-shinagawa, Shinagawa-ku, Tokyo, Japan
[3]Rakuten Institute of Technology Lab., Shinagawa Seaside Rakuten Tower, 4-12-3 Higashi-shinagawa, Shinagawa-ku, Tokyo, Japan
[4]Faculty of Policy Management, Keio University, Endo 5322, Fujisawa, Kanagawa, Japan
[5]Research Fellow, Rakuten Institute of Technology Lab., Shinagawa Seaside Rakuten Tower, 4-12-3 Higashi-shinagawa, Shinagawa-ku, Tokyo, Japan

Theme: Economy, Communications and Management

Introduction

In this research, we visualize and analyze the co-purchase networks of DVDs in order to understand the hidden relationship of respective DVDs in the Japanese online market. In the network, a node represents a DVD or a DVD-box, which means a set of DVDs, and the nodes are connected only if they are purchased by same customer. The source data of the network is collected by the store "Rakuten Books", which is one of the biggest online stores in Japan. The data we visualized are an extraction of 30,000 people from customers of the online store at random. Note that this research was done as an analysis by Rakuten Institute of Technology, and the data do not include any personal information.

Results

From this research, we visualize the relationships of the DVDs, not only a series of products. We also found the products classification by customers' choices. In addition, we draw the map of the products in the online market.

For visualizing the network, we try two types of connection method: "full connection" and "sequential connection". In the former method, all the nodes which user bought connect each other. In the latter method, nodes connect as the sequential order of user bought. It means that an undirected graph is generated by the former method and a direct graph is generated by the latter.

In the co-purchase networks of DVDs, for example in the case of "full connection", there are 3,118 nodes and 21,234 links, and the network has two large components, some middle-size components, and a lot of small clusters or pairs of products. One of the large components is formed by cartoon films and SF films, and the other component consists of various products, which are connected by very famous films like a Hollywood films. This means that some famous DVDs connect a genre to another genre, like a bridge between genres.